HEVC Enhancement using Content-based Local QP Selection

Fan Zhang[a] and David R. Bull[a]

[a]Department of Electrical and Electronic Engineering, University of Bristol, Bristol, BS8 1UB, United Kingdom

ABSTRACT

Inspired by recent advances in objective video quality assessment, this paper proposes a novel, local quantisation parameter (QP) determination approach for perceptual video compression, based on the experimental results of a QP selection test. This method has been fully integrated into the High Efficiency Video Coding (HEVC) reference codec for intra coding, which predicts coding tree unit (CTU) level QPs to achieve optimised rate quality performance. The proposed approach consistently shows bitrate savings based on perceptual quality metrics and Bjontegaard delta measurements, with minimal complexity increase over the original codec.

Keywords: P erceptual video compression, quantisation parameter selection, HEVC

1. INTRODUCTION

The greater demand for higher quality, more immersive video content is currently the primary driver for the development of internet, broadcasting and surveillance technologies. This significantly increases the required bandwidth, and challenges compression technologies.

For most applications, the objective of video compression is, under certain bitrate constraints, to provide optimum perceptual quality rather than to minimise the absolute distortion between compressed frames and the originals. The most recent standardised video codec, HEVC [1], represents an evolution of the conventional waveform coding framework, which is based on enhanced transformation, quantisation, and rate-distortion optimisation (RDO). In contrast, perceptual video compression algorithms are typically based on advanced computer vision and signal processing techniques, including texture analysis, warping and synthesis [2–5], and image inpainting [6].

One of the most important components of a perceptual video codec, is the quality assessment model. The last two decades have seen the development of numerous perception-based image and video quality assessment methods, which exploit different characteristics of the human visual system (HVS), including: just noticeable distortion (JND) [7, 8] methods [9], contrast sensitivity [10] based quality metrics [11], similarity measures [12, 13], assessment methods based on spatial/temporal information [14], and quality metrics inspired by the near/supra threshold perceptual strategy [11, 15, 16]. These have driven the improvement of various video coding tools, for example: luminance JND-based transformation [17] and quantisation [18, 19], and rate quality optimisation with SSIM [20].

Recently, Zhang and Bull proposed a perceptual video quality metric (PVM) [21, 22], which simulates the HVS perception process by adaptively combining texture masking based noticeable distortion and blurring artefacts, showing superior correlation performance with subjective opinions on a wide range of test video databases. In this work, spatial and temporal texture masking was shown to dominate in video quality perception, especially at low distortion levels (also known as the near threshold range) [22].

Inspired by PVM, this paper investigates the rate quality performance of HEVC intra coding (All Intra configuration) on sequences with mixed video content, and identifies the benefit of using dynamic local QP values. The results show a consistent correlation between spatio-temporal texture masks in PVM and optimum local QP values. This has been integrated into HEVC for intra coding with little additional complexity, determining the best QP value for every CTU. The integrated codec was evaluated on HEVC test sequences, and provides consistent bitrate savings, assessed by perceptual video quality metrics PVM and VQM [23], over video clips with various content and resolutions.

The rest of this paper is organised as follows: Section 2 describes the experiment on local QP selection, and presents the results and respective analysis. Based on this, Section 3 proposes a content-based QP selection

method. The compression results of the proposed approach are then given in Section 4. Finally, Section 5 provides conclusions.

2. A LOCAL QP SELECTION EXPERIMENT

Quantisation parameters (QPs) are employed in HEVC and H.264/AVC encoders (ranging between 0 and 51), to control the quantisation process of transform coefficients [24]. HEVC not only supports constant QP values for one sequence or frame, but also allows dynamic local QP variation within the range (± 7) for CTUs and Coding Units (CUs). The latter feature is mainly used in conjunction with rate control applications.

If a perceptual metric is used to measure video quality, then the resulting rate-quality characteristics of the codec will, in most cases, differ to when MSE or PSNR is used. This is primarily due to the influence of masking effects associated with the HVS. We therefore hypothesise that, if local QP values are selected according to content type, then this could result in improved rate-quality performance.

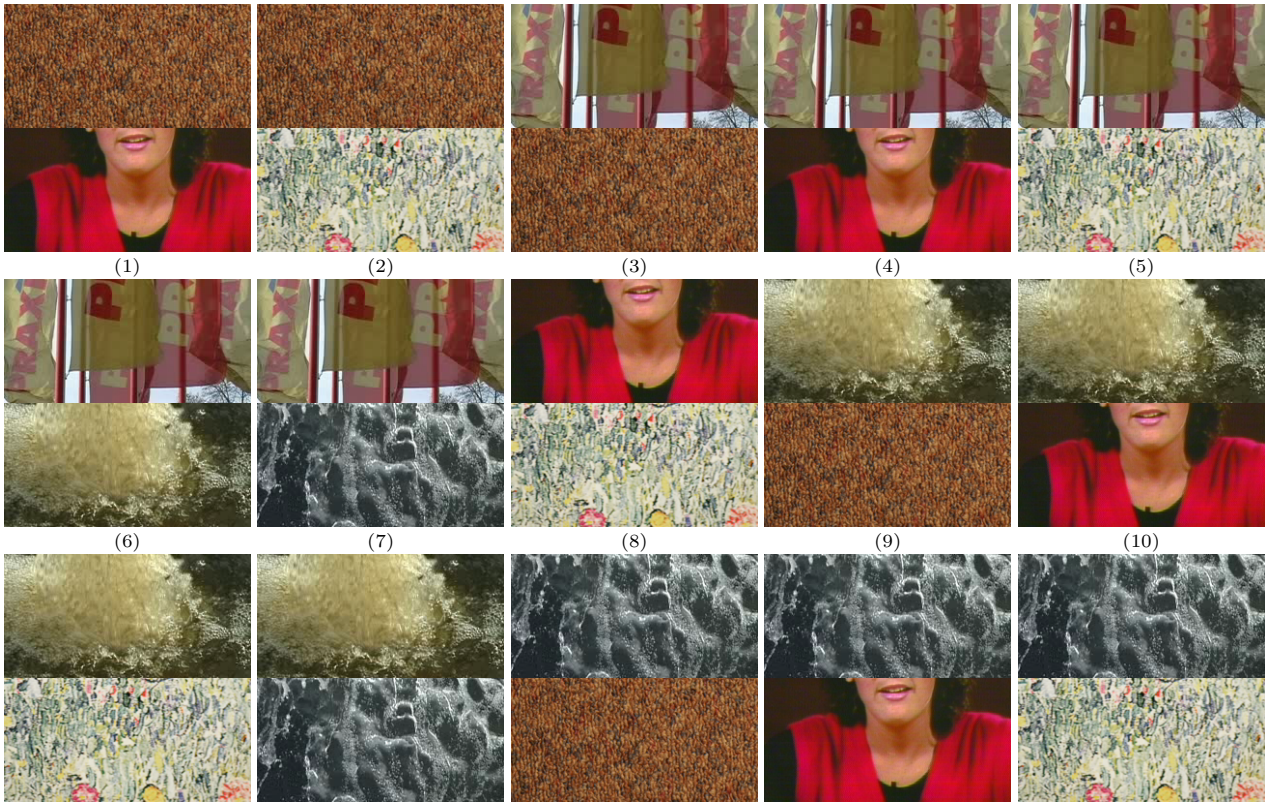

Figure 1: Test sequences used in the local QP selection experiment.

To prove this conjecture, a local QP selection experiment was conducted using the HEVC reference codec (HM 16.4) in the All Intra configuration (main profile). In order to simplify the experiment, fifteen artificial test sequences (YUV 4:2:0, 256×256, 100 frames) were generated by vertically combining two different types of material. The raw video clips are originally from the DynTex [25] and BVI Texture [26] databases, or are standard test sequences. All these sequences are shown and indexed in Fig. 1.

In this experiment, the rate quality performance using constant QP values are compared to the performance when different local QPs are used. Fifteen ΔQP test values are used, where ΔQP indicates the difference in QP between the top or bottom subframes ($\mathrm{QP}_{\mathrm{sub}}$) and the whole frame ($\mathrm{QP}_{\mathrm{frm}}$). Frame level $\mathrm{QP}_{\mathrm{frm}} = \{22, 27, 32, 37 \text{ and } 42\}$ are tested here.

$$\Delta\mathrm{QP} = \mathrm{QP}_{\mathrm{sub}} - \mathrm{QP}_{\mathrm{frm}}, \tag{1}$$

$$\Delta QP \in \{0, \pm 1, \pm 2, \pm 3, \ \ldots \ , \pm 7\}. \tag{2}$$

All CTUs in the test region (either the top or bottom half of each frame) use QP_{sub} values for the whole sequence, while the other half uses QP_{frm}. This generates thirty groups of rate quality results for each sequence at each test frame level QP, - fifteen for testing the top half, and fifteen for the bottom half.

The optimum QP differences ΔQP_{opt} are identified for all test sequences (for both top and bottom sub content) and frame level QPs. This is based on the corresponding overall rate distortion performance for all frames using both PSNR and PVM measurements, benchmarked against that using constant frame level QP values for all CTUs. Fig. 2 shows the comparison between ΔQP_{opt} and the corresponding frame level QPs.

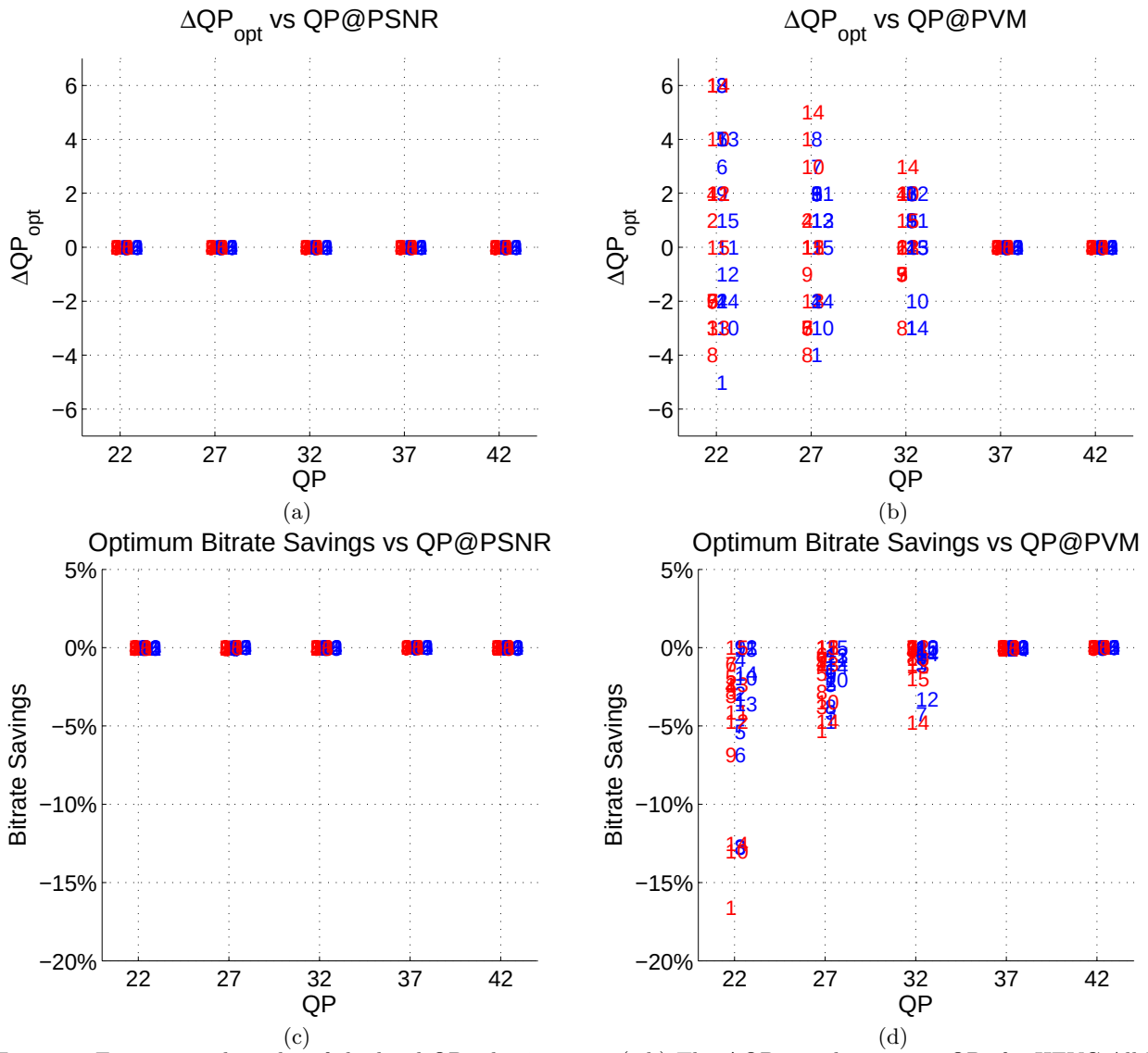

Figure 2: Experimental results of the local QP selection test. (a,b) The ΔQP_{opt} values versus QPs for HEVC All Intra configuration based on PSNR and PVM respectively. (c,d) Bitrate savings with corresponding ΔQP_{opt} values compared to HEVC using constant QPs, based on PSNR and PVM. The position of each number represents the ΔQP_{opt} value or bit rate saving for that sequence at a certain QP. Blue numbers indicate that the bottom half of frames is tested, while red numbers are used when the top half is tested. Negative values for bitrate savings represent better compression performance.

It can be observed that $\Delta QP_{opt} = 0$ for all test sequences if PSNR is used as the quality metric, which

leads to no bitrate savings for all cases. This indicates that frame level QPs are the optimum selection for all CTUs in terms of rate-PSNR performance. However when the perceptual quality metric, PVM, is employed, $\Delta\mathrm{QP}_{\mathrm{opt}}$ values vary according to the nature of the test content. Positive numbers occur when test content is more textured (spatial or temporal) than the other half, while negative values correspond to regions with lower texture content. The respective bitrate savings are significant, up to -15%, but are content dependent.

In order to predict optimum local QPs, we employ the spatial and temporal combined masks in PVM [22]. These have been effectively utilised for video quality assessment, and have contributed in feature analysis for characterising video content in [26, 27].

A spatial mask, $\mathrm{SM}(x,y)$, is defined as the maximum of six high frequency DT-CWT [28] subband coefficient magnitudes, $\left|B_o^j(x,y)\right|$, at the first level of decomposition of a original video frame (luminance only):

$$\mathrm{SM}(x,y) = \max\left\{\left|B_o^j(x,y)\right|, j = 1, 2, \ \ldots\ , 6\right\}. \tag{3}$$

A temporal mask is designed to characterise dynamic textures, and is based on approximated second derivatives of motion vectors:

$$\mathrm{TM}(x_b,y_b) = \max_{p=\pm1,\pm2}\{\frac{1}{|p|}\cdot \mathrm{SD}_p(x_b,y_b)\}. \tag{4}$$

To ensure consistency, motion estimation is applied between the current frame and its neighbouring frames (four neighbouring frames are used in this case - two in front of the current frame and two behind). The subscript 'p' indicates the displacement of the reference frame used, which can assume values of ±1 or ±2 from the current position. The combined second derivative $\mathrm{SD}(x_b,y_b)$ is defined as follows:

$$\mathrm{SD}_p(x_b,y_b) = \left\|\mathbf{SDX}_p(x_b,y_b)\right\|_2 + \left\|\mathbf{SDY}_p(x_b,y_b)\right\|_2. \tag{5}$$

$$\mathbf{SDX}(x_b,y_b) = \mathbf{MV}(x_b-1,y_b) + \mathbf{MV}(x_b+1,y_b) - 2\mathbf{MV}(x_b,y_b), \tag{6}$$

$$\mathbf{SDY}(x_b,y_b) = \mathbf{MV}(x_b,y_b-1) + \mathbf{MV}(x_b,y_b+1) - 2\mathbf{MV}(x_b,y_b). \tag{7}$$

Here, motion vectors $\mathbf{MV}(x_b,y_b)$ are obtained by applying an 8×8 block motion estimation on original frames, and the block level mask TM is further interpolated to the same size as SM.

Spatial and temporal masks are finally merged using (8):

$$M(x,y) = \max\{\rho_{\mathrm{sm}}\cdot\mathrm{SM}(x,y), \rho_{\mathrm{tm}}\cdot\mathrm{TM}(x,y)\}. \tag{8}$$

where ρ_{sm} and ρ_{tm} are empirically obtained from the LIVE video database [29], with constant values 2.2 and 0.25 respectively*. More details about spatial and temporal masks could be found in [22].

In this experiment, average values (sequence level) of combined masks for the whole frame and test regions are calculated as M_{frm} and M_{sub}, and their difference ΔM is then obtained to predict optimum local QPs.

$$\Delta M = M_{\mathrm{sub}} - M_{\mathrm{frm}}. \tag{9}$$

Fig. 3 shows the relationship between ΔM and $\Delta\mathrm{QP}_{\mathrm{opt}}$ for all test sequences and QPs. It can be seen that ΔM is directly proportional to $\Delta\mathrm{QP}_{\mathrm{opt}}$ for all five test QPs, and that their ratio follows a piecewise function in terms of the frame level QP, as illustrated in the last subfigure of Fig. 3. This relationship can be summarised as follows.

$$\Delta\mathrm{QP}_{\mathrm{opt}} = \begin{cases} (-0.066\mathrm{QP} + 2.531)\Delta M, & \mathrm{QP} < 38 \\ 0, & \mathrm{QP} \geq 38 \end{cases}. \tag{10}$$

*It is noted that the parameter values used here are different from those in [22]. This is because the training database [30] we have used in [22] contains interlaced content. In this paper, for progressive videos, we have trained the model using the same method in [22] based on the LIVE video database.

3. PROPOSED ALGORITHM

Based on the experimental results presented above, a content-based CTU level QP selection method has been developed, which predicts optimum local QPs using the local and global statistics of texture masks of PVM. Considering the high complexity of motion estimation, we omit the temporal mask in (8), and only use the spatial mask for prediction. This simplification has been validated in [21].

Before encoding a frame, F_i, with frame level quantisation parameter QP_i, the texture mask is calculated from the luminance channel of the original frame following (3) and (11).

$$M(x,y) = \rho_{\mathrm{sm}} \cdot M_s(x,y). \tag{11}$$

The average value of the mask for the given frame is denoted as M_i. When a coding tree unit $\mathrm{CTU}_{n,i}$ in frame F_i is about to be encoded, the mean of mask values for this CTU, $M_{n,i}$, is obtained and compared to M_i. The difference between them, $\Delta M_{n,i} = M_{n,i} - M_i$, is utilised to estimate the best QP, $\mathrm{QP}_{n,i}$, for this CTU.

$$\mathrm{QP}_{n,i} = \mathrm{QP}_i + \begin{cases} (-0.066\mathrm{QP} + 2.531)\Delta M_{n,i}, & \mathrm{QP} < 38 \\ 0, & \mathrm{QP} \geq 38 \end{cases}. \tag{12}$$

Here QP_i is the pre-determined frame level QP value. It should be also noted that due to the configuration of HEVC, the difference between $\mathrm{QP}_{n,i}$ and QP_i has been constrained within the range $[-7,\ +7]$.

4. RESULTS AND DISCUSSION

The proposed content-based local QP selection method has been integrated into the HEVC reference codec (HM 16.4), and was fully tested for All Intra configuration (main profile) over twenty one 8-bit JCT-VC recommended test sequences [31]. We follow the same test conditions as in [31], which employ frame level QPs from 22 to 37 with an interval of 5.

The compression performance of this approach is compared with the original HEVC codec, assessed by two perceptual video quality metrics - VQM and PVM [23]. The former is a commonly used video quality metric, standardised by ANSI and included in two ITU Recommendations. PVM [22], as described in Section 1, offers superior correlations with subjective quality scores compared to many existing video quality metrics. The rate quality results over all frames are based on the Bjontegaard delta approach [32]. Table 1 summaries the BD-savings using both quality metrics for all test sequences[†].

It can be observed that the proposed local QP selection method always performs better than the anchor codec, with BD-rate savings up to 3.5% and 3.8% for PVM and VQM respectively. The extent of this improvement depends on content type - sequences with homogeneous content achieve less improvement than those with spatially different content.

Finally, the computational complexity of the proposed method has been evaluated. In the context of HEVC, the extra time consumed using our approach was, on average, 4%. This figure was obtained using an Intel Core i7-2600 CPU @3.40GHz PC platform. The increased complexity is mainly due to the computation of the spatial masks.

5. CONCLUSIONS

In this paper, a novel content-based local QP determination approach has been presented for HEVC to improve rate quality performance. Inspired by the experimental results of a QP selection test, the texture mask model in PVM is employed to estimate optimum CTU QP values. This method has been integrated into HEVC reference codec for intra coding, and offers consistent bitrate savings over all the HEVC test sequences assessed by perceptual quality metrics. Future work will focus on the extension of this work to Low Delay and Random Access configurations.

[†]It is noted that standard VQM implementation [33] is not able to support 2560×1600 resolution video input, so we have not provided results for these two sequences.

Table 1: Summary of the compression results. Negative numbers indicate better compression efficiency.

Class	Sequence	BD-Rate (PVM)	BD-Rate (VQM)
A (2560×1600)	PeopleOnStreet	-0.4%	n/a
	Traffic	-0.8%	n/a
	Overall	-0.6%	n/a
B (1920×1080)	BQTerrace	-3.5%	-1.8%
	BasketballDrive	-0.8%	-0.8%
	Cactus	-1.9%	-3.3%
	Kimono1	-0.1%	-0.4%
	ParkScene	-0.6%	-0.9%
	Overall	-1.4%	-1.5%
C (832×480)	BQMall	-1.8%	-2.8%
	BasketballDrill	-3.4%	-2.2%
	PartyScene	-2.2%	-2.0%
	RaceHorses	-1.9%	-3.7%
	Overall	-2.3%	-2.7%
D (416×240)	BQSquare	-1.3%	-1.1%
	BasketballPass	-1.5%	-2.1%
	BlowingBubbles	-1.0%	-1.4%
	RaceHorses	-1.1%	-2.5%
	Overall	-1.2%	-1.8%
E (1280×720)	FourPeople	-1.1%	-1.1%
	Johnny	-0.9%	-0.1%
	KristenAndSara	-2.2%	-2.7%
	vidyo1	-0.8%	-1.6%
	vidyo3	-1.8%	-3.8%
	vidyo4	-1.8%	-2.8%
	Overall	-1.5%	-2.0%

6. REFERENCES

[1] G. J. Sullivan, J. R. Ohm, W. J. Han, and T. Wiegand, "Overview of the high efficiency video coding (HEVC) standard," *IEEE Trans. on Circuits and System for Video Technology*, vol. 22, no. 12, pp. 1649–1668, 2012.

[2] P. Ndjiki-Nya, T. Hinz, and T. Wiegand, "Generic and robust video coding with texture analysis and synthesis," in *Proc. IEEE Int Conf. on Multimedia & Expo*. IEEE, 2007, pp. 1447–1450.

[3] F. Zhang and D. R. Bull, "A parametric framework for video compression using region-based texture models," *IEEE Journal of Selected Topics in Signal Processing*, vol. 5, no. 7, pp. 1378–1392, 2011.

[4] M. Bosch, F. Zhu, and E. J. Delp, "Segmentation-based video compression using texture and motion models," *IEEE Journal of Selected Topics in Signal Processing*, vol. 5, no. 7, pp. 1277–1281, 2011.

[5] J. Balle, A. Stojanovic, and J. R. Ohm, "Models for static and dynamic texture synthesis in image and video compression," *IEEE Journal of Selected Topics in Signal Processing*, vol. 5, no. 7, pp. 1353–1365, 2011.

[6] C. Zhu, X. Sun, F. Wu, and H. Li, "Video coding with spatio-temporal texture synthesis and edge-based inpainting," in *Proc. IEEE Int Conf. on Multimedia & Expo*. IEEE, 2008, pp. 813–816.

[7] H. L. F. von Helmholtz, *Handbook of Physiological Optics*, Voss, Hamburg and Leipzig, Germany, 1st edition, 1896.

[8] H. R. Blackwell, "Luminance difference threshold," in *Handbook of Sensory Physiology*, D. Jameson and L. M. Murvich, Eds., pp. 78–101. Springer-Verlag, New York, 1972.

[9] X. Zhang, W. Lin, and P Xue, "Improved estimation for just-noticeable visual distortion," *Signal Processing*, vol. 84, no. 4, pp. 795–808, 2005.

[10] D. H. Kelly, "Motion and vision. II. stabilized spatio-temporal threshold surface," *Journal of Optical Society of America*, vol. 69, no. 10, pp. 1340–1349, 1979.

[11] D. Chandler and S. Hemami, "VSNR: A wavelet-based visual signal-to-noise ratio for natural images," *IEEE Trans. on Image Processing*, vol. 16, no. 9, pp. 2284–2298, 2007.

[12] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Trans. on Image Processing*, vol. 13, pp. 600–612, 2004.

[13] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multi-scale structural similarity for image quality assessment," in *Proc. Asilomar Conference on Signals, Systems and Computers*. IEEE, 2003, vol. 2, p. 1398.

[14] K. Seshadrinathan and A. C. Bovik, "Motion tuned spatio-temporal quailty assessment of natural videos," *IEEE Trans. on Image Processing*, vol. 19, pp. 335–350, 2010.

[15] D. M. Chandler, "Seven chanllenges in image quality assessment: Past present, and future research," *ISRN Signal Processing*, vol. 2013, 2013.

[16] P. V. Vu, C. T. Vu, and D. M. Chandler, "A spatiotemporal most-apparent-distortion model for video quality assessment," in *Proc. IEEE Int Conf. on Image Processing*. IEEE, 2011, pp. 2505–2508.

[17] A. J. Ahumada and H. A. Peterson, "Luminance-model-based DCT quantization for color image compression," in *Proc. SPIE: Human Vision, Visual Process.* SPIE, 1993, vol. 1666.

[18] M. Naccari and F. Pereira, "Advanced h.264/avc-based perceptual video coding: Architecture, tools and assessment," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 21, no. 6, pp. 766–782, 2011.

[19] Y. Zhang, M. Naccari, D. Agrafiotis, M. Mrak, and D. Bull, "High dynamic range video compression exploiting luminance masking," *IEEE Trans. on Circuits and System for Video Technology*, vol. 26, no. 5, pp. 950–964, 2016.

[20] S. S. Channappayya, A. C. Bovik, and R. W. Heath, "Rate bounds on SSIM index of quantised images," *IEEE Trans. on Image Processing*, vol. 617, no. 9, pp. 1624–1639, 2008.

[21] F. Zhang and D. Bull, "Quality assessment method for perceptual video compression," in *Proc. IEEE Int Conf. on Image Processing*, 2013, pp. 39–43.

[22] F. Zhang and D. Bull, "A perception-based hybrid model for video quality assessment," *IEEE Trans. on Circuits and System for Video Technology*, in press.

[23] M. H. Pinson and S. Wolf, "A new standardized method for objectively measuring video quality," vol. 50, pp. 312–322, 2004.

[24] D. R. Bull, *Communicating pictures: A course in Image and Video Coding*, Academic Press, 2014.

[25] R. Péteri, S. Fazekas, and M. J. Huiskes, "DynTex: a comprehensive database of dynamic textures," *Pattern Recognition Letters*, vol. 31, pp. 1627–1632, 2010, http://projects.cwi.nl/dyntex/.

[26] M. A. Papadopoulos, F. Zhang, D. Agrafiotis, and D. R. Bull, "A video texture database for perceptual compression and quality assessment," in *Proc. IEEE Int Conf. on Image Processing*, 2015.

[27] F. M. Moss, K. Wang, F. Zhang, R. Baddeley, and D. R. Bull, "On the optimal presentation duration for subjective video quality assessment," *IEEE Trans. on Circuits and System for Video Technology*, in press.

[28] N. G. Kingsbury, "Complex wavelets for shift invariant analysis and filtering of signals," *Journal of Applied and Computational Harmonic Analysis*, vol. 10, no. 3, pp. 234–253, 2001.

[29] K. Seshadrinathan, R. Soundararajan, A. C. Bovik, and L. K. Cormack, "Study of subjective and objective quality assessment of video," *IEEE Trans. on Image Processing*, vol. 19, pp. 335–350, 2010.

[30] Video Quality Experts Group, "Final report from the video quality experts group on the validation of objective quailty metrics for video quality assessment.," Tech. Rep., VQEG, 2000.

[31] J. R. Ohm, G. J. Sulivan, H. Schwarz, T. K. Tan, and T. Wiegand, "Comparison of the coding efficiency of video coding sstandard - including High Efficiency Video Coding (HEVC)," *IEEE Trans. on Circuits and System for Video Technology*, vol. 22, no. 12, pp. 1669–1684, 2012.

[32] G. Bjontegaard, "Calculation of average PSNR differences between RD-curves," Tech. Rep., VCEG-M33 Meeting, Austin. TX, April 2001.

[33] ITS, "Video quality metric (VQM) software," http://www.its.bldrdoc.gov/resources/video-quality-

research/software.aspx.

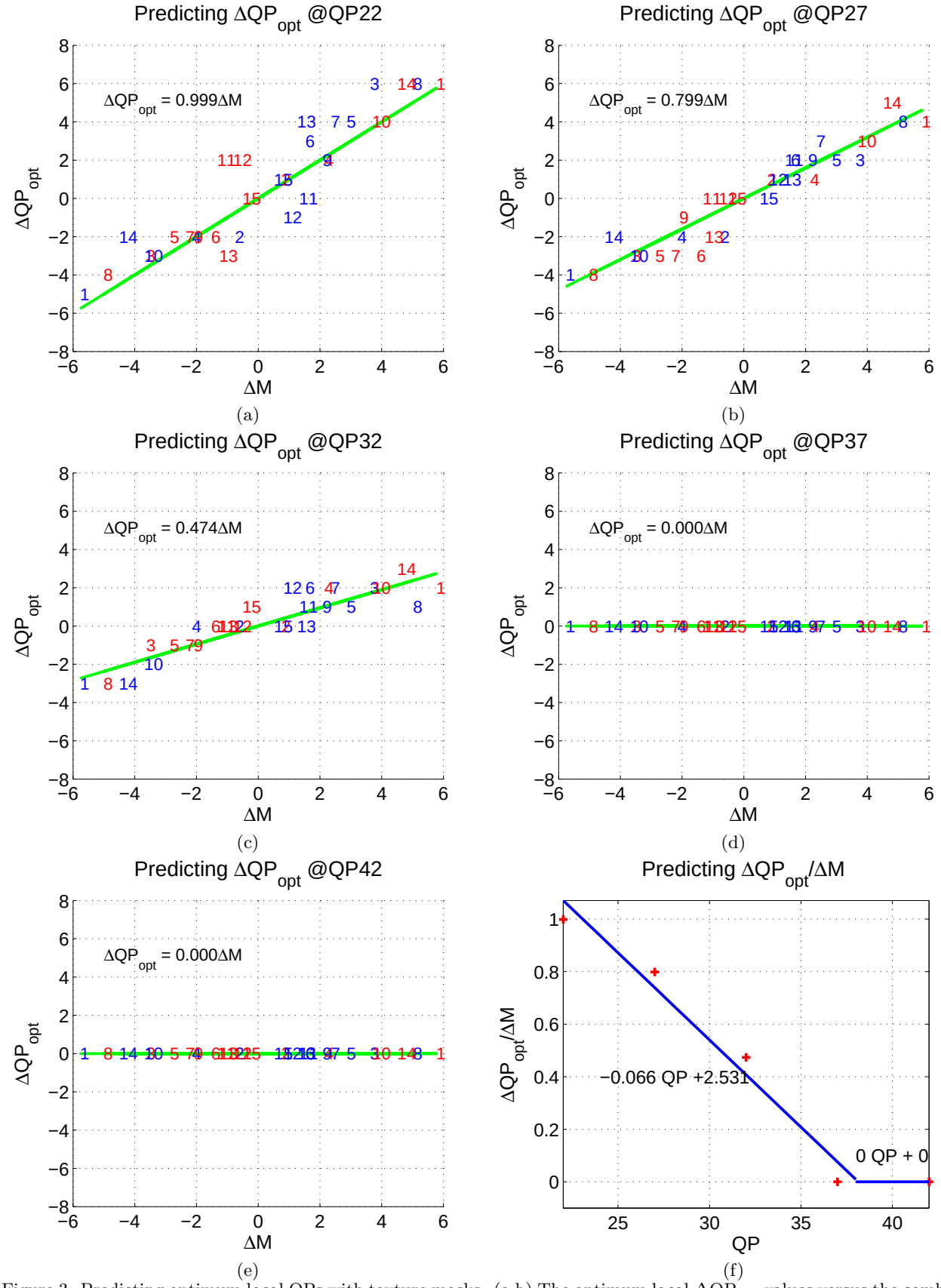

Figure 3: Predicting optimum local QPs with texture masks. (a,b) The optimum local ΔQP_{opt} values versus the combined texture masks M of corresponding sequences for HEVC All Intra configuration at QP 22 and 27 respectively. (c,d) ΔQP_{opt} versus, M, for QP 32 and 37. (e) ΔQP_{opt} versus M for QP 42. (f) $\Delta QP/M$ versus frame level QPs. The position of each number represents the ΔQP_{opt} value for that sequence with certain ΔM value. Blue indicates that the bottom half of frames is tested, while red indicates the top half.